

# Binary equality sets are generated by two words

Štěpán Holub

Faculty of Mathematics and Physics, Charles University  
186 75 Praha 8, Sokolovská 83, Czech Republic  
`holub@karlin.mff.cuni.cz`

## Abstract

We show that the equality set  $\text{Eq}(g, h)$  of two non-periodic binary morphisms  $g, h : A^* \rightarrow \Sigma^*$  is generated by at most two words. If the rank of  $\text{Eq}(g, h) = \{\alpha, \beta\}^*$  is two, then  $\alpha$  and  $\beta$  start (and end) with different letters.

This in particular implies that any binary language has a test set of cardinality at most two.

## About this version

This is a revised version of my paper published (with the same title) in *Journal of Algebra* 259 (2003), 1–42.

A nucleus of the paper was a part of my Ph.D. thesis supervised by Aleš Drápal ([10]). The proof was completed during the postdoctoral stay in Turku granted by Turku Centre for Computer Science (TUCS). I am grateful especially to Juhani Karhumäki for making that stay possible. When writing the paper I discussed the topic with Vesa Halava, Tero Harju, Juhani Karhumäki and Juha Kortelainen.

After the publication, I received comments from Elena (Petre) Czeizler, Markku Laine and Václav Flaška. The present version was carefully read by Jiří Sýkora. I am indebted to all of them for their effort, their comments and suggestions.

The most important difficulty discovered was Lemma 29, which does not hold as it stays in the published text. The corrected formulation given in the present version is the one from an early draft of the paper. Before the publication, I decided to use a stronger claim, which is in fact never needed in the paper, and which, as it turned out, is fallacious. Elena pointed out some difficulties in the proof, and Markku found a counterexample, making it clear that the stronger claim cannot be rescued.

The present version was written in October 2007 and August 2012. The material is partly reorganized, terminology is revised and most proofs rewritten. I hope that the text is now substantially more readable than the journal version,

which is in some places excessively complicated and discouraging. On the other hand, there are no new discoveries and the overall argument remains the same.

## 1 Introduction

Binary equality language, i.e., the set on which two binary morphisms agree, is the most simple non-trivial example of an equality language, the notion of which was introduced in [9]. Equality languages in general play an important role in formal language theory. For a survey and bibliography see [6, Section 5].

In the binary case, the morphisms are defined on a monoid generated by two letters. It was for the first time extensively studied by K. Čulík II and J. Karhumäki in [3]. There, the main claim of our work was conjectured, viz. that a binary equality language is generated by at most two words as soon as at least one of the morphisms is non-periodic (or, equivalently, injective). An important step towards the proof of the conjecture was made in [4] where the following partial characterization was obtained.

**Theorem 1.** *The equality set of two binary morphisms  $g, h : A^* \rightarrow \Sigma^*$ , where  $A = \{a, b\}$ , has the following structure:*

(A) *If  $g$  and  $h$  are periodic, then either  $\text{Eq}(g, h) = \{\varepsilon\}$  or*

$$\text{Eq}(g, h) = \{\varepsilon\} \bigcup \{\alpha \in A^+ \mid \frac{|\alpha|_a}{|\alpha|_b} = k\}$$

*for some  $k \geq 0$  or  $k = \infty$ .*

(B) *If exactly one morphism is periodic, then*

$$\text{Eq}(g, h) = \alpha^*$$

*for some word  $\alpha \in A^*$ .*

(C) *If both  $g$  and  $h$  are non-periodic, then either*

$$\text{Eq}(g, h) = \{\alpha, \beta\}^*$$

*for some words  $\alpha, \beta \in A^*$ , or*

$$\text{Eq}(g, h) = (\alpha\gamma^*\beta)^*$$

*for some words  $\alpha, \beta, \gamma \in A^+$ .*

The question remained open whether the second possibility of case (C), contradicting the conjecture, can actually occur. In the present paper we show that the answer is negative and, moreover, if  $\alpha$  and  $\beta$  are both nonempty, they start (and end) with different letters. This is formulated in the following main theorem.

**Theorem 2.** *Let  $g, h : A^* \rightarrow \Sigma^*$  be non-periodic binary morphisms. Let  $\alpha$  and  $\beta$ , with  $\alpha \neq \beta$ , be nonempty minimal elements of  $\text{Eq}(g, h)$ . Then*

$$\text{pref}_1(\alpha) \neq \text{pref}_1(\beta) \quad \text{and} \quad \text{suff}_1(\alpha) \neq \text{suff}_1(\beta).$$

As a trivial consequence we have a solution of the original question.

**Theorem 3.** *Equality language of two nonperiodic binary morphisms is generated by at most two words.*

I am not aware of any way how to prove Theorem 3 not using Theorem 2.

*Remark.* Later, in [7], it has been shown that the equality sets generated by two words have a precise form. Namely, the following theorem holds true.

**Theorem.** *Let  $g$  and  $h$  be distinct nonperiodic binary morphisms such that  $\text{Eq}(g, h)$  is generated by two words. Then there is a positive integer  $i$  such that*

$$\text{Eq}(g, h) = \{a^i b, ba^i\}^*,$$

*up to renaming of the letters.*

The proof is based on Theorem 2.

A closely related problem is the size of a test set for binary languages. Indeed, if two morphisms agree on a language, it must be a subset of their equality language. In [4], it is shown that all binary languages have a three element test set. Our result allows to cut down this bound to two. Let us remark that this improvement is not a simple consequence of the fact that the equality language is generated by two words — the difference in the first (or last) letter is a necessary ingredient.

## 2 Preliminaries

In this section, we fix our notation and recall some basic facts. For a reference and unproved claims see [2] or [8]. If  $\Sigma$  is an alphabet, then let  $\Sigma^*$  be the free monoid, and  $\Sigma^+$  the free semigroup generated by  $\Sigma$ . The empty word is denoted by  $\varepsilon$ . Any subset of  $\Sigma^*$  is called a *language*. Let  $A$  denote the two-letter alphabet  $\{a, b\}$ .

The length of the word is denoted by  $|u|$ , and  $|u|_x$  denotes the number of occurrences of the letter  $x$  in  $u$ . A *prefix* of  $u$  is any word  $v \in \Sigma^*$  such that there exists a word  $v' \in \Sigma^*$  with  $u = vv'$ . The set of all prefixes of  $u$  is denoted by  $\text{pref}(u)$ . A prefix  $v$  of  $u$  is *proper* if  $v \neq \varepsilon$  and  $v \neq u$ . Similarly, *suffix* and *proper suffix* are defined. The set of all suffixes of  $u$  is denoted by  $\text{suff}(u)$ . The first (the last resp.) letter of a nonempty word  $u$  is denoted by  $\text{pref}_1(u)$  ( $\text{suff}_1(u)$  resp.). A word  $v$  is called a *factor* of  $u$  if there exist words  $w, w' \in \Sigma^*$  such that  $u = wvw'$ .

If  $v \in \text{pref}(u)$  or  $u \in \text{pref}(v)$ , then we say that  $u$  and  $v$  are *prefix-comparable* (or simply *comparable*). The maximal common prefix of words  $u$  and  $v$  is denoted by  $u \wedge v$ . If  $u$  and  $v$  are words, then the maximal *u-prefix* of  $v$  is the

maximal prefix of  $v$  that is also a prefix of  $u^i$  for some  $i$ . Analogously, we define the maximal  $u$ -suffix of  $v$ . We say that two words are *suffix-comparable* if one is a suffix of the other.

Positive powers  $u^n$  of a word are defined as usual, with  $u^0 = \varepsilon$ . We shall sometimes use also negative powers and work with elements of the free group, in order to simplify notation. This should not cause any confusion. For example, if  $u$  and  $v$  are comparable, then we shall write  $u^{-1}v$  even if  $v$  is a proper prefix of  $u$ . In such a case, when  $u = vw$ , we have  $u^{-1}v = w^{-1}$ .

We shall define regular languages by regular expressions in a standard way. In particular, the language  $\{u^i \mid i \geq 1\}$  is denoted by  $u^+$  and  $u^* = u^+ \cup \{\varepsilon\}$ . We say that  $v$  is a prefix (suffix, factor resp.) of  $u^+$  if  $v$  is a prefix (suffix, factor resp.) of  $u^i$  for some  $i \geq 1$ .

A nonempty word  $u$  is called *primitive* if and only if  $u = v^n$  implies  $u = v$ . The *primitive root* of a nonempty word  $u$  is the (uniquely given) primitive word  $r$  such that  $u \in r^+$ . Words  $u$  and  $v$  are called *conjugate* if  $u = ww'$  and  $v = w'w$  for some words  $w$  and  $w'$ .

If we speak about minimality or maximality of some element, the implicit ordering is the prefix one, i.e.,  $v \leq u$  if and only if  $v \in \text{pref}(u)$ , and  $v < u$  if moreover  $v \neq u$ . (While by the *shortest* word we mean the word with the smallest length.)

Let  $u \in \Sigma^+$  be a word  $u = l_1 l_2 \dots l_d$ , with  $d = |u|$  and  $l_i \in \Sigma$ . Then the *reversal* of the word  $u$ , denoted by  $\overline{u}$ , is obtained by inverting the order of the letters, viz.

$$\overline{u} = l_d l_{d-1} \dots l_1.$$

Let  $g$  be an arbitrary morphism. The *reversal* of  $g$  is the morphism denoted by  $\overline{g}$ , which has the same range and domain as  $g$ , and is defined by

$$\overline{g}(x) = \overline{g(x)},$$

for each  $x \in \Sigma$ . Note that in general  $\overline{g}(u)$  does not equal to  $g(\overline{u})$  nor to  $\overline{g(u)}$ . Instead

$$\overline{g}(\overline{u}) = \overline{g(u)}.$$

All concepts and reasonings regarding prefixes are valid analogously for suffixes, reversals considered. We shall often use the fact.

A morphism  $g$  defined on  $\Sigma$  is called *erasing* if  $g(x)$  is empty for some  $x \in \Sigma$ . A morphism  $g$  is *periodic* if there is a word  $t$  such that  $g(x) \in t^*$ , for all words  $x$  (or, equivalently, all letters  $x$ ). Note that a binary morphism is periodic as soon as it is erasing.

Let  $S = T^+$  be a subsemigroup of  $\Sigma^+$  generated by a set  $T$ . The *rank* of  $T$  is the cardinality of the minimal set generating  $S$ . We can write

$$\text{rank}(T) = \text{rank}(S) = \text{Card}(S \setminus S \cdot S).$$

By the rank of a monoid  $M$  we mean the rank of the semigroup  $M \setminus \{\epsilon\}$ .

It is a well known fact that for each set  $M \subset \Sigma^+$  there exists the smallest free subsemigroup of  $\Sigma^+$  containing  $M$  and called its *free hull*. A set generating

a free semigroup is called a *code*. If any two distinct elements of a code are neither prefix nor suffix comparable, the set is called a *bifix code*.

The *equality set* of two morphisms  $g, h : \Delta^* \rightarrow \Sigma^*$  is defined by

$$\text{Eq}(g, h) = \{u \in \Delta^* \mid g(u) = h(u)\}.$$

It is easy to verify that the set  $\text{Eq}(g, h)$  is a free submonoid of  $\Delta^*$  generated by the set of its minimal elements

$$\text{eq}(g, h) = \text{Eq}(g, h) \setminus (\text{Eq}(g, h) \setminus \{\varepsilon\})^2 \setminus \{\varepsilon\}.$$

Note that  $\text{eq}(g, h)$  is a bifix code.

Let  $g : A^* \rightarrow \Sigma^*$  be a nonperiodic binary morphism. By  $z_g$  we denote the maximal common prefix of  $g(ab)$  and  $g(ba)$ , i.e.

$$z_g = g(ab) \wedge g(ba).$$

Since  $g$  is nonperiodic, we have  $|z_g| < |g(a)| + |g(b)|$  by Lemma 5 below. If  $\text{pref}_1(g(a)) \neq \text{pref}_1(g(b))$ , i.e.  $z_g = \varepsilon$ , we say that  $g$  is *marked*.

Similarly we define  $\underline{z}_g$  as the maximal common suffix of  $g(ab)$  and  $g(ba)$ . Note that

$$\underline{z}_g = \overline{g(ab)} \wedge \overline{g(ba)} = \overline{z_{\bar{g}}}$$

and  $\underline{z}_g = \varepsilon$  is equivalent to  $\bar{g}$  being marked.

Cartesian product  $\Delta^* \times \Delta^*$  is the set of ordered pairs  $(u, v)$  of words. It can be seen as a monoid with operation of catenation defined by  $(u, v)(u', v') = (uu', vv')$ , with the unit  $(\varepsilon, \varepsilon)$ . Such a monoid is obviously not free, it is even not isomorphic to a submonoid of a free monoid.

Let  $g, h : \Delta^* \rightarrow \Sigma^*$  be two morphisms. The subset of  $\Delta^* \times \Delta^*$  denoted by  $\mathbb{C}(g, h)$  and defined by

$$\mathbb{C}(g, h) = \{(u, v) \mid g(u) = h(v)\}$$

will be called the *coincidence set* of morphisms  $g$  and  $h$ . It is generated by the set

$$\mathbf{c}(g, h) = \mathbb{C}(g, h) \setminus (\mathbb{C}(g, h) \setminus \{(\varepsilon, \varepsilon)\})^2 \setminus \{(\varepsilon, \varepsilon)\}.$$

Any pair  $(u, v) \in \mathbb{C}(g, h)$  can be uniquely factorized into minimal pairs  $(u_i, v_i)$  satisfying  $g(u_i) = h(v_i)$ . This is formulated in the following lemma.

**Lemma 4.** *Let  $g$  and  $h$  be non-erasing morphisms. Then  $\mathbb{C}(g, h)$  is, as a submonoid of  $\Delta^* \times \Delta^*$ , freely generated by  $\mathbf{c}(g, h)$ . Moreover, the set  $\mathbf{c}(g, h)$  is a bifix code.*

Note that  $(u, u)$  is an element of  $\mathbb{C}(g, h)$  for each  $u \in \text{Eq}(g, h)$ , and  $\text{Eq}(g, h)$  is given uniquely by  $\mathbb{C}(g, h)$  as

$$\text{Eq}(g, h) = \{u \mid (u, u) \in \mathbb{C}(g, h)\}.$$

We present several combinatorial lemmas for future (often implicit) reference. Following three lemmas are part of the folklore.

**Lemma 5.** *The words  $u$  and  $v$  commute if and only if they have the same primitive root.*

**Lemma 6** (Periodicity Lemma). *Let  $u^+$  and  $v^+$  have a common prefix of length  $|u| + |v|$ . Then the words  $u$  and  $v$  commute.*

We shall often use the following lemma. It is based on the well known fact that a primitive word  $t$  cannot satisfy equality  $tt = utv$ , with  $u$  and  $v$  nonempty.

**Lemma 7.** (A) *Let  $ww = uwv$ . Then  $u$ ,  $v$  and  $w$  commute.*

(B) *Let  $uw$  be a prefix of  $w^+$ . Then  $u$  and  $w$  commute.*

(C) *Let  $sw$  be a factor of  $w^+$ . Then  $s$  is a suffix of  $w^+$ .*

(D) *Let  $uw$  be a suffix of  $w^+$  and let  $w$  be a prefix of  $uw$ . Then  $u$  and  $w$  commute.*

(E) *Let  $u_1, u_2, w, w' \in \Sigma^+$  be words such that  $w'$  and  $w$  are conjugate,  $|u_1| \leq |u_2|$ , and the words  $u_1w'$ ,  $u_2w'$  are prefixes of  $w^+$ . Then  $u_1$  is a suffix of  $u_2$  and  $u_2u_1^{-1}$  commutes with  $w$ .*

One more lemma, which is easy to prove:

**Lemma 8.** *Let  $g : A^* \rightarrow A^*$  be a marked morphism and let  $u, v \in A^*$ . Then  $g(u \wedge v) = g(u) \wedge g(v)$ .*

The following nice lemma is a key fact about binary morphisms.

**Lemma 9.** *Let  $X = \{x, y\} \subseteq \Sigma^+$  be a nonperiodic set (i.e.  $xy \neq yx$ ). Let  $u \in xX^*$ ,  $v \in yX^*$  be words such that  $|u|, |v| \geq |xy \wedge yx|$ . Then  $u \wedge v = xy \wedge yx$ .*

The proof is not difficult (see [2], p. 348). The lemma immediately implies that for a nonperiodic binary morphism  $h$  and an arbitrary word  $u \in A^+$  long enough, the word  $z_h$  is a prefix of  $h(u)$  and the  $(|z_h| + 1)$ -th letter of  $h(u)$  indicates the first letter of  $u$ . For any  $u, v \in A^*$  we have

$$z_h = h(au)z_h \wedge h(bv)z_h. \quad (1)$$

It is now easy to see that the morphism  $h_m$  such that

$$h_m(u) = z_h^{-1}h(u)z_h, \quad (2)$$

$u \in A$ , is well defined. Moreover, it is marked, and the equality (2) holds for any  $u \in A^*$ . We shall call it the *marked version* of  $h$ .

**N.B.** The case  $g = h$  is trivial. Throughout the paper we shall implicitly suppose  $g \neq h$ .

### 3 Principal morphisms

In this section we show that at least one of the morphisms  $g$  and  $h$  can be supposed to be marked. As we shall see, this will make our research more convenient. The goal is achieved by choosing a suitable target alphabet.

**Definition 10.** We say that an (unordered) pair of binary morphisms  $g, h : A^* \rightarrow \Sigma^*$  is *principal* if the target alphabet  $\Sigma$  generates the free hull of the set  $\{g(a), g(b), h(a), h(b)\}$ .

The previous definition reflects the use of the term “principal morphism” in literature (see for example [8], p. 170). The advantages of principal morphisms stem from the following important property.

**Lemma 11.** *Let  $X$  be a finite subset of  $\Sigma^*$  and let  $Y$  be the minimal generating set of the free hull of  $X$ . Then for each element  $y \in Y$  there is a word  $x \in X$  such that  $y$  is a prefix (suffix resp.) of  $x$ .*

For the proof see [1], Lemma 3.1. For our purpose, note the following immediate corollary.

**Corollary 12.** *Let  $X$  be a finite subset of  $\Sigma^*$  such that  $\Sigma$  is the base of the free hull of  $X$ . Then*

$$\Sigma = \{\text{pref}_1(u) \mid u \in X\} = \{\text{suff}_1(u) \mid u \in X\}.$$

It is quite intuitive that choosing the minimal generating set of the free hull as the target alphabet has no influence on the coincidence set of the morphisms. The following lemma is formulated for binary morphisms, but it can be trivially extended to any domain alphabet.

**Lemma 13.** *Let  $g_1, h_1$  be morphisms  $A^* \rightarrow \Sigma^*$ . Then there is a principal pair of morphisms  $g, h$  such that*

$$\mathbb{C}(g, h) = \mathbb{C}(g_1, h_1).$$

*Moreover, if  $g_1 (h_1, \overline{g_1}, \overline{h_1} \text{ resp.})$  is marked, then such is also  $g (h, \overline{g}, \overline{h} \text{ resp.})$ .*

*Proof.* Let  $F \subset \Sigma^*$  be the free hull of the set  $\{g_1(a), g_1(b), h_1(a), h_1(b)\}$  and let  $C$  be an alphabet whose cardinality equals the rank of  $F$ . Then  $C^*$  and  $F$  are isomorphic since they are both free monoids of the same rank; let  $\varphi : C^* \rightarrow F$  be an isomorphism. Define morphisms  $g, h : A^* \rightarrow C^*$  by

$$g = \varphi^{-1} \circ g_1, \quad h = \varphi^{-1} \circ h_1.$$

$$\begin{array}{ccc} & & F \\ & \nearrow^{g_1, h_1} & \uparrow \varphi \\ A^* & & \\ & \searrow_{g, h} & \downarrow \varphi^{-1} \\ & & C^* \end{array}$$

Then  $(g, h)$  is a principal pair of morphisms, the above diagram commutes, and  $\mathbb{C}(g, h) = \mathbb{C}(g_1, h_1)$ . The rest is obvious.  $\square$

The previous lemma shows that we can always, without loss of generality, suppose that the pair we work with is principal. We can now prove that this brings about markedness of one of the morphisms.

**Lemma 14.** *Let  $g, h$  be nonperiodic principal morphisms, with  $\text{eq}(g, h)$  non-empty. Then at least one of the morphisms  $g, h$  is marked, and at least one of the morphisms  $\bar{g}, \bar{h}$  is marked.*

*Proof.* Suppose that none of the morphisms is marked, therefore

$$\text{pref}_1(g(a)) = \text{pref}_1(g(b)), \quad \text{pref}_1(h(a)) = \text{pref}_1(h(b)).$$

Let  $x$  be a first letter of a word  $u \in \text{Eq}(g, h)$ . Then

$$\text{pref}_1(g(x)) = \text{pref}_1(h(x)),$$

and Corollary 12 implies that the morphisms are periodic, a contradiction.

Obviously, the morphisms  $\bar{g}, \bar{h}$  are also principal, since the concept of the free hull is preserved under the reversal symmetry. This concludes the proof.  $\square$

## 4 The block structure of the coincidence set

In this section, we study the structure of the equality set of nonperiodic morphisms and their relation to the coincidence set. The previous section justifies why we shall always suppose that  $g$  is marked.

Let  $u, v \in \Sigma^*$  be words such that  $g(u)$  and  $h(v)$  are comparable. Then the word  $h(v)^{-1}g(v)$  is called an *overflow* (the overflow may be a “negative” word if  $g(v)$  is a prefix of  $h(v)$ ). Following lemmas show that the possibility to lengthen the words  $u, v$  to words  $u', v'$  such that  $g(u') = h(v')$  is very restricted. Namely, the overflow  $z_h$  is the only one admitting two different continuations.

**Lemma 15.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let  $u, v \in A^*$  be words such that  $g(u)$  and  $h(v)$  are comparable and let*

$$g(u) \neq h(v)z_h.$$

*Let  $u_1, u_2, v_1, v_2 \in A^+$  be words such that*

$$g(uu_1) = h(vv_1), \quad g(uu_2) = h(vv_2).$$

*Then*

- $\text{pref}_1(u_1) = \text{pref}_1(u_2)$ , if  $|g(u)| - |h(v)| < |z_h|$ ;
- $\text{pref}_1(v_1) = \text{pref}_1(v_2)$ , if  $|g(u)| - |h(v)| > |z_h|$ .

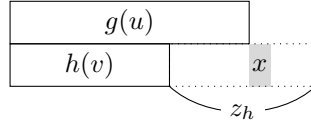


*Proof.* If  $u_1, u_2, v_1$  and  $v_2$  satisfy the conditions of the lemma, then the same conditions are satisfied also by the words  $u_1uu_1, u_2uu_2, v_1vv_1$  and  $v_2vv_2$  resp. Hence we can suppose that each of the words  $u_1, u_2, v_1, v_2$  is longer than  $z_h$ . Consider three cases.

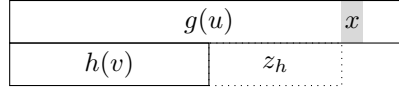
1. First suppose that  $|g(u)| < |h(v)| + |z_h|$ . By (1),  $h(v)z_h$  is a prefix of both  $h(vv_1)$  and  $h(vv_2)$  and

$$\text{pref}_1(g(u_1)) = \text{pref}_1(g(u_2)) = \text{pref}_1(g(u)^{-1}h(v)z_h) = x.$$

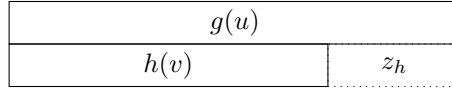
Since  $g$  is a marked morphism, this implies that  $\text{pref}_1(u_1) = \text{pref}_1(u_2)$ .



2. Suppose on the other hand that  $|g(u)| > |h(v)| + |z_h|$ . Then  $h(v_1), h(v_2)$  have the common prefix longer than  $z_h$  and  $\text{pref}_1(v_1) = \text{pref}_1(v_2)$  is determined by the letter  $x = \text{pref}_1((h(v)z_h)^{-1}g(u))$ .



3. If  $|g(u)| = |h(v)| + |z_h|$ , then, clearly,  $g(u) = h(v)z_h$ .



□

Previous lemma yields the following property.

**Lemma 16.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let  $(c, d)$  and  $(c', d')$  be distinct elements of  $\mathbb{C}(g, h)$ , and suppose that  $c$  and  $c'$  are not comparable. Put*

$$u = c \wedge c', \quad v = d \wedge d'.$$

*Then*

$$g(u) = h(v)z_h.$$

*Proof.* We have  $c = uu_1$  and  $c' = uu_2$  where  $u_1, u_2 \in A^+$  and  $\text{pref}_1(u_1) \neq \text{pref}_1(u_2)$ .

If  $d$  and  $d'$  are not comparable, then  $d = vv_1$  and  $d' = vv_2$  with  $v_1, v_2 \in A^+$  and  $\text{pref}_1(v_1) \neq \text{pref}_1(v_2)$ , and the claim follows from Lemma 15.

If  $d$  and  $d'$  are comparable, then  $|g(u)| - |h(v)| < 0 \leq |z_h|$ . Since

$$g(uu_1c) = h(vv_1d), \quad g(uu_2c') = h(vv_2d')$$

with  $u_1c, u_2c', v_1d, v_2d' \in A^+$ , Lemma 15 yields a contradiction with  $\text{pref}_1(u_1) \neq \text{pref}_1(u_2)$ . □

*Example 17.* The previous corollary does not hold without the condition that  $c$  and  $c'$  are not comparable. Consider morphisms

$$\begin{aligned} g(a) &= a, & g(b) &= b, \\ h(a) &= a, & h(b) &= aab. \end{aligned}$$

Then  $(c, d) = (a, a)$ ,  $(c', d') = (aab, b)$ ,  $z_h = aa$ , and

$$g(c \wedge c') = g(a) = a \neq aa = h(\varepsilon)z_h = h(d \wedge d')z_h.$$

The ground for the characterization of the coincidence set is the following lemma.

**Lemma 18.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let the words  $e, f \in A^+$  satisfy following conditions:*

- (i)  $z_h g(e) = h(f)z_h$
- (ii) *The words  $e, f$  are minimal, i.e.: If  $u \leq e$ ,  $v \leq f$  and  $z_h g(u) = h(v)z_h$ , then either  $u = v = \varepsilon$  or  $u = e$  and  $v = f$ .*

*Then, given the first letter of  $e$  or the first letter of  $f$ , the words  $e$  and  $f$  are determined uniquely.*

*Proof.* Suppose  $e, f$  and  $e', f'$  satisfy (i) and (ii), and  $\text{pref}_1(e) = \text{pref}_1(e')$ . Put  $c = e \wedge e'$ ,  $d = f \wedge f'$ . Since  $g$  is a marked morphism, we have

$$z_h g(e) \wedge z_h g(e') = z_h g(c) \tag{3}$$

by Lemma 8. From (1) we deduce

$$h(f)z_h \wedge h(f')z_h = h(d)z_h. \tag{4}$$

Since  $z_h g(e) = h(f)z_h$  and  $z_h g(e') = h(f')z_h$ , the equalities (3), (4) yield

$$z_h g(c) = h(d)z_h.$$

Since  $c$  is nonempty, we deduce from (ii) that  $c = e = e'$  and  $d = f = f'$ . Similarly if  $\text{pref}_1(f) = \text{pref}_1(f')$ .  $\square$

This implies the following lemma.

**Lemma 19.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked.*

- (A) *The rank of  $\mathbb{C}(g, h_m)$  is at most two.*
- (B) *If the rank of  $\mathbb{C}(g, h_m)$  is two and  $\mathbf{c}(g, h_m) = \{(e, f), (e', f')\}$ , then*

$$\begin{aligned} \text{pref}_1(e) &\neq \text{pref}_1(e') \\ \text{pref}_1(f) &\neq \text{pref}_1(f'). \end{aligned}$$

*Proof.* Recall that  $h_{\mathbf{m}}(u) = z_h^{-1}h(u)z_h$  to see that

$$\mathbb{C}(g, h_{\mathbf{m}}) = \{(u, v) \in A^* \times A^* \mid z_h g(u) = h(v)z_h\}.$$

The rest is a consequence of Lemma 18.  $\square$

Note that  $(e, f) \in \mathbf{c}(g, h_{\mathbf{m}})$  is just another formulation of the fact that  $e, f$  are minimal words satisfying  $z_h g(e) = h(f)z_h$ , which are exactly conditions of Lemma 18. The pairs  $(e, f)$  and  $(e', f')$  are often called *blocks* of  $g$  and  $h$ .

The question on the structure of the equality set  $\text{Eq}(g, h)$  can be seen as a special case of the above considerations. If conditions

$$u = v, \quad u_1 = v_1, \quad u_2 = v_2, \quad c = d, \quad c' = d', \quad e = f, \quad e' = f',$$

are added, then we get the following modifications of Lemma 15, Lemma 16, Lemma 18 and Lemma 19 with analogous proofs, which we omit.

**Lemma 20.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let  $u \in A^*$  be a word such that  $g(u)$  and  $h(u)$  are comparable, and*

$$g(u) \neq h(u)z_h.$$

*Let  $u_1, u_2 \in A^+$  be words such that*

$$\begin{aligned} g(uu_1) &= h(uu_1), \\ g(uu_2) &= h(uu_2). \end{aligned}$$

*Then  $\text{pref}_1(u_1) = \text{pref}_1(u_2)$ .*

**Lemma 21.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let  $c$  and  $c'$  be incomparable elements of  $\text{Eq}(g, h)$ . Put  $u = c \wedge c'$ . Then*

$$g(u) = h(u)z_h.$$

**Lemma 22.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked. Let the word  $e \in A^+$  satisfy following conditions:*

- (i)  $z_h g(e) = h(e)z_h$
- (ii) *The word  $e$  is minimal, i.e.: If  $e_1$  is a prefix of  $e$  and  $z_h g(e_1) = h(e_1)z_h$ , then  $e_1 = \varepsilon$  or  $e_1 = e$ .*

*Then the word  $e$  is determined uniquely by its first letter.*

**Lemma 23.** *Let  $g$  and  $h$  be binary morphisms, and let  $g$  be marked.*

- (A) *The rank of  $\text{Eq}(g, h_{\mathbf{m}})$  is at most two.*
- (B) *If the rank of  $\text{Eq}(g, h_{\mathbf{m}})$  is two and  $\text{eq}(g, h_{\mathbf{m}}) = \{e, e'\}$ , then*

$$\text{pref}_1(e) \neq \text{pref}_1(e').$$

Note that the previous lemma proves Theorem 2 for morphisms, which are marked from both sides. In the rest of the paper we show that this is essentially the only situation in which the equality set can have rank greater than one.

Marked morphisms are in general much easier to deal with. That's why it is convenient to work with principal pairs, where one of the morphisms, say  $g$ , is marked. Moreover, it is always possible to use the marked version  $h_m$  instead of  $h$  to get a marked pair, and thus a better insight into the coincidence set of  $g$  and  $h$ .

The block structure of the coincidence set of marked morphisms leads to an important concept of *successor morphisms* introduced first in [5]. Consider marked morphisms  $g$  and  $h$  such that  $\mathbf{c}(g, h)$  consists of two blocks  $(e, f)$  and  $(e', f')$ . Let  $w$  be an element of  $\text{Eq}(g, h)$ . The equality  $g(w) = h(w)$  can be uniquely split into a sequence of blocks. This means that  $w$  is an element of  $\{e, e'\}^+$ , and in the same time an element of  $\{f, f'\}^+$ . It is now natural to define the successor morphisms  $(g_1, h_1)$  by

$$\begin{cases} g_1(a) = e, \\ g_1(b) = e', \end{cases} \quad \begin{cases} h_1(a) = f, \\ h_1(b) = f', \end{cases} \quad (5)$$

and to formulate the previous considerations by the following lemma.

**Lemma 24.** *Let  $g, h$  be marked morphisms such that*

$$\mathbf{c}(g, h) = \{(e, f), (e', f')\}.$$

*Then the morphisms  $g_1, h_1$  defined by (5) are marked. If  $w \in \text{Eq}(g, h)$ , then there is a unique word  $w_1 \in \text{Eq}(g_1, h_1)$  such that*

$$g_1(w_1) = h_1(w_1) = w.$$

*Proof.* The morphisms  $g_1$  and  $h_1$  are marked by Lemma 19. The existence and uniqueness of the word  $w_1$  follows from  $(w, w) \in \mathbb{C}(g, h)$ , and from Lemma 4.  $\square$

## 5 The counterexample and its structure

We now have all necessary ingredients for the proof of our main claim, Theorem 2. The course of the prove will be essentially by contradiction. We shall assume that there exists a counterexample to the claim, and gradually show that such an assumption is contradictory.

We first formulate what is understood as a counterexample.

**Definition 25.** We say that a pair of morphisms  $(g, h)$  is a *counterexample* if

- (a) The rank of  $\text{Eq}(g, h)$  is at least two;
- (b)  $g$  is marked and  $h$  is not marked;

$$(c) |g(a)| > |h(a)| \text{ and } |g(b)| < |h(b)|.$$

The third condition takes advantage of the symmetry of letters  $a$  and  $b$ . Note that the strict inequalities do not harm generality, since  $|g(a)| = |h(a)|$  or  $|g(b)| = |h(b)|$  would imply  $g = h$ . Since the letters  $a$  and  $b$  are not interchangeable anymore, we shall sometimes need the morphism  $\pi$  defined by  $\pi(a) = b$  and  $\pi(b) = a$ .

The following lemma yields basic information about the structure of the equality set of a counterexample.

**Lemma 26.** *Let  $(g, h)$  be a counterexample. Then there exist nonempty words  $\sigma$ ,  $\nu_a$  and  $\nu_b$  such that  $|\sigma|_a \geq 1$ ,*

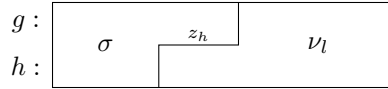
$$\text{pref}_1(\nu_a) = a, \quad \text{pref}_1(\nu_b) = b,$$

*the words  $\sigma\nu_a$ ,  $\sigma\nu_b$  are two distinct elements of  $\text{eq}(g, h)$  and*

$$g(\sigma) = h(\sigma)z_h, \quad (6)$$

$$z_h g(\nu_a) = h(\nu_a), \quad (7)$$

$$z_h g(\nu_b) = h(\nu_b). \quad (8)$$



*Proof.* Let  $u$  and  $v$  be two distinct elements of  $\text{eq}(g, h)$ . Note that  $u$  and  $v$  are not comparable, and put  $\sigma = u \wedge v$ ,  $u_1 = \sigma^{-1}u$  and  $v_1 = \sigma^{-1}v$ . Clearly,  $\text{pref}_1(u_1) \neq \text{pref}_1(v_1)$  and the choice of  $\nu_a$  and  $\nu_b$  is now obvious. The equalities (6), (7) and (8) are yielded by Lemma 21, and  $|\sigma|_a \geq 1$  follows from  $|g(b)| < |h(b)|$ .  $\square$

The equalities (6), (7) and (8) are of a special importance in the proof. They represent two points, where the structure of a counterexample is well defined, and which therefore yield information for a combinatorial analysis.

The following lemma makes sure that the counterexample defined above deserves its name.

**Lemma 27.** *Let  $g_1$  and  $h_1$  be nonperiodic binary morphisms such that  $\text{eq}(g_1, h_1)$  contains two elements  $\alpha$  and  $\beta$  with the same first letter. Then there is a counterexample  $(g, h)$  such that  $\text{Eq}(g, h) = \text{Eq}(g_1, h_1)$ .*

*Moreover, if  $\overline{g_1}$  ( $\overline{h_1}$  resp.) is marked, then also  $\overline{g}$  ( $\overline{h}$  resp.) is marked.*

*Proof.* Lemma 13 yields principal morphisms  $g$  and  $h$  such that  $\text{Eq}(g, h) = \text{Eq}(g_1, h_1)$ . By Lemma 14 and by the symmetry of  $g$  and  $h$ , we can suppose that  $g$  is marked. Similarly, by the symmetry of  $a$  and  $b$ , we can suppose that the condition (c) of Definition 25 is satisfied. In order to see that  $(g, h)$  is a counterexample, it remains to show that  $h$  is not marked. If  $h$  is marked, then both morphisms are marked, and  $\text{pref}_1(\alpha) \neq \text{pref}_1(\beta)$  by Lemma 23, contrary to the assumption.

Markedness of reversals is conserved by Lemma 13.  $\square$

The further strategy is to show that there is no counterexample. We shall divide the investigation into several stages.

## 6 When $z_h$ commutes

In this section we investigate two special situations, in which  $z_h$  commutes with one of the image words. We show that those situations lead to a contradiction. We start with a technical lemma, which will be the core of the proof. In the original version of this paper the claim had the following strong form:

**Lemma.** Let  $g, h : A^* \rightarrow A^*$  be two marked morphisms. Let  $u, u', v$  and  $v' \in A^*$  be words, and  $s, r, q$  positive integers such that

$$g(a^s bu) = h(a^s bu'), \quad g(a^r bv) = h(a^q bv').$$

Then  $s = r = q$ .

However, as Markku Laine pointed out by constructing an example, this claim does not hold. The example is as follows.

*Example 28.* Let

$$\begin{aligned} g(a) &= a^2 b^2, & h(a) &= a, \\ g(b) &= b, & h(b) &= b^2. \end{aligned}$$

Then

$$g(a^2 b^2) = h(a^2 b a^2 b b) = a^2 b^2 a^2 b^4,$$

and

$$g(ab^2) = h(a^2 b^2) = a^2 b^4.$$

We therefore present a bit weaker version, which fits the purpose of this paper.

**Lemma 29.** Let  $g, h : A^* \rightarrow A^*$  be two marked morphisms. Let  $u$  and  $v \in A^*$  be words, and  $s, r, q$  positive integers such that

$$g(a^s bu) = h(a^s bu), \tag{9}$$

$$g(a^r bv) = h(a^q bv). \tag{10}$$

Then  $s = r = q$ .

*Proof.* Recall that we suppose  $g \neq h$ . (Obviously, only  $r = q$  is forced if  $g = h$ .) Let  $g$  and  $h$  be morphisms satisfying assumptions, and suppose that  $s = r = q$  does not hold. Assume, moreover, that  $g$  and  $h$  are chosen such that the length of  $a^s bu$  is the smallest possible. We show that  $a^s bu$  can be shortened, and hence obtain a contradiction.

We first prove that  $g(a)$  and  $h(a)$  do not commute. Suppose for a while that  $|g(a)| > |h(a)|$ , and that  $t$  is the common primitive root of  $g(a)$  and  $h(a)$ . From

(9), we deduce that  $h(b)$  is comparable with  $h(a^s)^{-1}g(a^s)$ , which is an element of  $t^+$ . That is a contradiction with  $h$  being marked. Similarly if  $|g(a)| < |h(a)|$ . (Clearly,  $g(a) = h(a)$  implies  $g = h$ .)

Let us continue the proof of the lemma. Lemma 8 applied once to  $g$  and once to  $h$  gives

$$g(a^s bu \wedge a^r bv) = g(a^s bu) \wedge g(a^r bv) = h(a^s bu) \wedge h(a^q bv) = h(a^s bu \wedge a^q bv). \quad (11)$$

1. If  $s \neq r$  and  $s \neq q$ , then (11) yields

$$g(a^i) = h(a^j),$$

with  $i = \min(s, r)$ ,  $j = \min(s, q)$ . Therefore the words  $g(a)$  and  $h(a)$  commute, a contradiction.

2. Suppose next, by symmetry,  $s = r$  and  $s \neq q$ . Put  $m = \min(s, q)$ . Equality (11) implies

$$g(a^s bw) = h(a^m), \quad (12)$$

where  $w = u \wedge v$ .

The set  $\mathbb{C}(g, h)$  contains elements  $(a^s bu, a^s bu)$  and  $(a^s bw, a^m)$ , whence the rank of  $\mathbb{C}(g, h)$  is two. Let  $(e, f)$  and  $(e', f')$  be the blocks of  $g$  and  $h$ , and let  $g_1, h_1$  be their successor morphisms defined by (5).

By symmetry, suppose that  $\text{pref}_1(f) = a$ . Equality (12) implies that there is a positive integer  $p$  such that  $f = a^p$ . Since  $g(a)$  and  $h(a)$  do not commute, we deduce that  $e \notin a^+$  and thus  $|e| > s$ . Since  $a^s bu$  and  $a^q bv$  are elements of  $\{f, f'\}^*$ , both  $s$  and  $q$  are multiples of  $p$ . Put

$$s_1 = \frac{s}{p}, \quad q_1 = \frac{q}{p},$$

and define words  $u_1$  and  $v_1$  by

$$\begin{aligned} g_1(u_1) &= a^s bu, & h_1(u_1) &= a^s bu, \\ g_1(v_1) &= a^s bv, & h_1(v_1) &= a^q bv. \end{aligned}$$

Since  $f = a^p$ , the words  $u_1$  and  $v_1$  can be factorized as

$$u_1 = a^{s_1} bu_2, \quad v_1 = a^{q_1} bv_2,$$

with  $u_2, v_2 \in A^*$ . Therefore

$$\begin{aligned} g_1(a^{s_1} bu_2) &= h_1(a^{s_1} bu_2) = a^s bu, \\ g_1(a^{q_1} bv_2) &= h_1(a^{q_1} bv_2) = a^q bv. \end{aligned}$$

Inequality  $s \neq q$  implies  $s_1 \neq q_1$ , and  $|e| > s$  yields  $|a^{s_1} bu_2| < |a^s bu|$ . This completes the proof.  $\square$

The following two claims exploit the previous lemma. The words  $\sigma$ ,  $\nu_a$  and  $\nu_b$  are as in Lemma 26.

**Claim 1.** *There is no counterexample such that  $z_h$  commutes with  $g(b)$  and  $\text{pref}_1(\sigma) = b$ .*

*Proof.* Suppose that  $(g, h)$  is such a counterexample, and let  $t$  be the common primitive root of  $z_h$  and  $g(b)$ . Let  $b^\ell$  be the maximal  $b$ -prefix of  $\sigma\nu_a$  and  $b^k$  be the maximal  $b$ -prefix of  $\nu_b\sigma$ . Then  $g(b)^\ell$  is the maximal  $t$ -prefix of  $g(\sigma\nu_a)$  and  $z_h g(b)^k$  is the maximal  $t$ -prefix of  $z_h g(\nu_b\sigma)$ .

Suppose that  $h(b)$  commutes with  $g(b)$ . Since  $|h(b)| > |g(b)|$ , the equality  $g(\sigma\nu_a) = h(\sigma\nu_a)$  implies that  $g(a)$  is comparable with  $g(b)^{-\ell} h(b)^\ell$ , a contradiction with  $g$  being marked. Therefore  $h(b)$  and  $g(b)$  do not commute.

By Lemma 7(B), the maximal  $t$ -prefix of  $h(b)z_h$  is shorter than  $|h(b)t|$ . This implies, by (1), that all words  $h(bu)$  long enough have the same maximal  $t$ -prefix. In particular, the maximal  $t$ -prefix of  $h(\sigma\nu_a)$  is the same as the maximal  $t$ -prefix of  $h(\nu_b\sigma)$ . From  $h(\sigma\nu_a) = g(\sigma\nu_a)$  and  $h(\nu_b\sigma)z_h = z_h g(\nu_b\sigma)$  we deduce that  $g(b)^\ell = z_h g(b)^k$  and  $k \neq \ell$ .

Put  $\sigma' = b^{-\ell}\sigma$  and note that  $\sigma'$  is nonempty since  $\sigma$  contains the letter  $a$ . Then

$$z_h g(b^k \sigma') = h(b^\ell \sigma') z_h, \quad z_h g(\nu_b \sigma) = h(\nu_b \sigma) z_h,$$

and Lemma 29, applied to morphisms  $h_m \circ \pi$  and  $g \circ \pi$ , yields a contradiction.  $\square$

**Claim 2.** *There is no counterexample such that  $\text{pref}_1(\sigma) = a$ ,  $z_h$  commutes with  $h(a)$ , and the common primitive root of  $z_h$  and  $h(a)$  is a suffix of  $g(a)$ .*

*Proof.* As in the previous proof, suppose that  $(g, h)$  satisfies assumptions of the claim and let  $t$  be the common primitive root of  $z_h$  and  $h(a)$ . Let  $a^\ell$  be the maximal  $a$ -prefix of  $\sigma\nu_b$ , and  $a^k$  be the maximal  $a$ -prefix of  $\nu_a\sigma$ .

First, suppose that  $g(a)$  commutes with  $h(a)$  and  $z_h$ . Since  $g$  is marked, the maximal  $t$ -prefix of  $z_h g(\nu_a\sigma)$  is  $z_h g(a)^k$ . From (1), we deduce that the word  $z_h$  is the maximal  $t$ -prefix of  $h(bu)z_h$  for any  $u$ . Hence the maximal  $t$ -prefix of  $h(\nu_a\sigma)z_h$  is  $h(a)^k z_h$ . The equality  $z_h g(\nu_a\sigma) = h(\nu_a\sigma)z_h$  now yields  $z_h g(a)^k = h(a)^k z_h$ , a contradiction with  $|g(a)| > |h(a)|$ . Therefore  $g(a)$  and  $h(a)$  do not commute.

Since, by assumption,  $t$  is a suffix of  $g(a)$ , Lemma 7(B) implies that the maximal  $t$ -prefix of  $g(\sigma\nu_b) = h(\sigma\nu_b)$  is equal to the maximal  $t$ -prefix of  $g(a)$ . Using (1) as above, we deduce from that this maximal  $t$ -prefix is equal to  $h(a)^\ell z_h$ . In this case, the equality  $z_h g(\nu_a\sigma) = h(\nu_a\sigma)z_h$  implies  $z_h h(a)^\ell z_h = h(a)^k z_h$  whence  $z_h = h(a)^{k-\ell}$ .

For  $\sigma' = a^{-\ell}\sigma$  we obtain

$$z_h g(a^\ell \sigma') = h(a^k \sigma') z_h, \quad z_h g(\nu_a \sigma) = h(\nu_a \sigma) z_h.$$

Since  $g(a)$  and  $h(a)$  do not commute, we deduce that  $\sigma \notin a^+$ , whence  $\text{pref}_1(\sigma') = b$  and morphisms  $h_m$ ,  $g$  satisfy assumptions of Lemma 29, a contradiction.  $\square$



## 7 Case: $\bar{g}$ is not marked

In this section we deal with the situation when  $\bar{g}$  is not marked. Note that then  $\bar{h}$  is marked by Lemma 14, and verify that  $(\bar{h} \circ \pi, \bar{g} \circ \pi)$  is also a counterexample. Recall that  $\pi$  exchanges letters  $a$  and  $b$ , and it is applied in order to satisfy the condition (c) of Definition 25. This allows to suppose

$$|\underline{z}_g| \geq |z_h|. \quad (13)$$

More precisely, if  $|\underline{z}_g| < |z_h|$ , then we consider  $(\bar{h} \circ \pi, \bar{g} \circ \pi)$ , instead of  $(g, h)$ .

The equality (1) applied to reversals implies that  $\underline{z}_g$  is a suffix of any  $g(u)$  long enough. Especially,

$$\underline{z}_g \text{ is a suffix of } g(a)^+, \quad (14)$$

$$\underline{z}_g \text{ is a suffix of } g(b)^+. \quad (15)$$

Since  $z_h$  is a suffix of  $g(\sigma)$ , which is suffix comparable with  $\underline{z}_g$ , we deduce from (13) that

$$z_h \in \text{suff}(\underline{z}_g). \quad (16)$$

$$\begin{array}{l} g : \boxed{\begin{array}{cc} & \underline{z}_g \\ \sigma & \end{array}} \\ h : \boxed{\begin{array}{cc} & \\ \sigma & z_h \end{array}} \end{array}$$

The following claim excludes the situation of this section.

**Claim 3.** *There is no counterexample with  $\bar{g}$  not marked.*

*Proof.* Suppose that  $(g, h)$  is such a counterexample.

1. Suppose first  $\text{pref}_1(\sigma) = a$ . The equality  $g(\sigma) = h(\sigma)z_h$  yields  $h(a) \in \text{pref}(g(a))$ , and  $z_h g(\nu_a) = h(\nu_a)$  implies that  $h(a)z_h$  is a prefix of  $z_h g(a)$ . Thus  $z_h h(a) = h(a)z_h$ .

Let  $t$  be the common primitive root of  $h(a)$  and  $z_h$ . From (16) we deduce that  $t$  is a suffix of  $\underline{z}_g$ , and (14) together with  $|g(a)| > |h(a)| \geq |t|$  yields that  $t$  is a suffix of  $g(a)$ . This is a contradiction with Claim 2.

2. Suppose then that  $\text{pref}_1(\sigma) = b$ . From (15) and (16) we deduce that  $z_h g(b)$  is a suffix of  $g(b)^+$ . Equalities  $g(\sigma) = h(\sigma)z_h$  and  $z_h g(\nu_b) = h(\nu_b)$  imply that  $g(b)$  is a prefix of  $h(b)$ , and that  $h(b)$  is comparable with  $z_h g(b)$  respectively. Therefore  $g(b)$  is a prefix of  $z_h g(b)$ , and Lemma 7(D) yields that  $g(b)$  and  $z_h$  commute, a contradiction with Claim 1.  $\square$

## 8 Case: $\bar{h}$ is not marked

In this subsection we consider the situation when  $\bar{g}$  is marked and  $\bar{h}$  is not. We shall not exclude this case directly. Instead we reduce it to the case when both  $\bar{g}$  and  $\bar{h}$  are marked.

To accomplish this plan we first we need a description of possible counterexample structure that is more precise than Lemma 26.

**Lemma 30.** *Let  $(g, h)$  be a counterexample. Then one of the following possibilities takes place.*

(A) *There exist words  $\sigma, \mu_a, \mu_b \in A^+$ , and  $\tau \in A^*$  such that*

$$\text{eq}(g, h) = \{\sigma\mu_a\tau, \sigma\mu_b\tau\},$$

*where*

$$\begin{aligned} z_h g(\mu_a) \underline{z}_h &= h(\mu_a), & g(\sigma) &= h(\sigma) z_h, \\ z_h g(\mu_b) \underline{z}_h &= h(\mu_b), & g(\tau) &= \underline{z}_h h(\tau), \end{aligned}$$

*and*

$$\text{pref}_1(\mu_a) = a, \quad \text{pref}_1(\mu_b) = b, \quad \text{suff}_1(\mu_a) \neq \text{suff}_1(\mu_b).$$

$$\begin{array}{l} g : \\ h : \end{array} \begin{array}{|c|c|c|c|c|} \hline & & & & \\ \hline \sigma & & \mu_x & & \tau \\ \hline & z_h & & \underline{z}_h & \\ \hline \end{array}$$

(B) *There exist words  $\zeta, \mu, \rho, \eta \in A^+$  such that*

$$\text{eq}(g, h) = \zeta(\rho\mu)^* \rho \eta = \zeta \rho (\mu\rho)^* \eta,$$

*and*

$$\begin{aligned} g(\zeta) \underline{z}_h &= h(\zeta), & z_h g(\mu) \underline{z}_h &= h(\mu), & \text{pref}_1(\mu) &\neq \text{pref}_1(\eta), \\ g(\rho) &= \underline{z}_h h(\rho) z_h, & z_h g(\eta) &= h(\eta), & \text{suff}_1(\mu) &\neq \text{suff}_1(\zeta). \end{aligned}$$

$$\begin{array}{l} g : \\ h : \end{array} \begin{array}{|c|c|c|c|c|c|c|} \hline & & & & & & \\ \hline \zeta & & \rho & & \mu & & \eta \\ \hline & \underline{z}_h & & z_h & & \underline{z}_h & z_h \\ \hline \end{array}$$

*Proof.* Let  $\alpha$  and  $\beta$  be two shortest elements of  $\text{eq}(g, h)$ . Put  $\sigma = \alpha \wedge \beta$ , and similarly let  $\tau$  be the longest common suffix of  $\alpha$  and  $\beta$ . By Lemma 21, applied first to  $g$  and  $h$ , and then to  $\bar{g}$  and  $\bar{h}$ , we have

$$g(\sigma) = h(\sigma) z_h, \tag{17}$$

$$g(\tau) = \underline{z}_h h(\tau). \tag{18}$$

Denote by  $v_0$  and  $v_1$  the words  $\sigma^{-1}\alpha$  and  $\sigma^{-1}\beta$ . Clearly,  $\text{pref}_1(v_0) \neq \text{pref}_1(v_1)$ .

1. First suppose that  $v_0$  and  $v_1$  are not suffix-comparable. Then with a suitable choice of  $i, j \in \{0, 1\}$  we have  $v_i = \mu_a\tau$ ,  $v_j = \mu_b\tau$ , and  $\text{pref}_1(\mu_\ell) = \ell$  for both  $\ell \in A$ .

Therefore  $\{\sigma\mu_a\tau, \sigma\mu_b\tau\} = \{\alpha, \beta\}$ . We show that  $\sigma$  is the unique prefix of  $\alpha$  ( $\beta$  resp.) satisfying (17).

Suppose first that  $\sigma_1\sigma_2 = \sigma$  and  $g(\sigma_1) = h(\sigma_1)z_h$ . Then it is easy to see that also  $\sigma_1v_i \in \text{Eq}(g, h)$ ,  $i = 0, 1$ , a contradiction with  $\alpha$  and  $\beta$  being the shortest elements of  $\text{eq}(g, h)$ .

Let then  $v_i = w_1w_2$ , for some  $i \in \{0, 1\}$ , and  $g(\sigma w_1) = h(\sigma w_1)z_h$ . Then  $\sigma w_2$  is an element of  $\text{Eq}(g, h)$ , which is shorter than  $\sigma v_i$ . Since  $\alpha$  and  $\beta$  are the shortest elements of  $\text{eq}(g, h)$ , it remains that  $\sigma w_2 = \sigma v_{1-i}$ . But then  $v_0$  and  $v_1$  are suffix-comparable, a contradiction.

We still have to show that the set  $\{\alpha, \beta\}$  generates whole  $\text{Eq}(g, h)$ . Suppose that  $w$  is an element of  $\text{Eq}(g, h)$  such that neither  $\alpha$ , nor  $\beta$  is a prefix of  $w$ , and consider words  $w_i = w \wedge \sigma v_i$ ,  $i = 0, 1$ . Lemma 21 implies that  $g(w_i) = h(w_i)z_h$ , for both  $i = 0, 1$ . It is easy to deduce that  $w_0$  and  $w_1$  cannot be both equal to  $\sigma$ , a contradiction with the previous paragraph. Consequently, we have the case (A).

2. Suppose now, by symmetry, that  $v_1 = uv_0$ . Then  $z_h g(u) = h(u)z_h$  and  $\sigma u^*v_0$  is a subset of  $\text{Eq}(g, h)$ . Moreover,  $\sigma$  and  $\sigma u$  are the only prefixes of  $\sigma uv_0$  satisfying (17). The proof is similar as above: any other prefix satisfying (17) allows to drop a part of the word, which contradicts the minimality of  $\alpha$  and  $\beta$ . We omit details. This, in particular, implies that  $u$  is not a suffix of  $\sigma$ .

We show that  $\sigma u^*v_0$  generates the whole equality set. Suppose the contrary, and let  $w$  be the shortest element of  $\text{eq}(g, h)$  that is not in  $\sigma u^*v_0$ . As above, the words  $w_0 = w \wedge \sigma v_0$  and  $w_1 = w \wedge \sigma uv_0$  satisfy  $g(w_i) = h(w_i)z_h$ ,  $i = 0, 1$ . Therefore  $w_0 = \sigma$ , by the previous paragraph. From  $\text{pref}_1(u) \neq \text{pref}_1(v_0)$ , one obtains that  $w_1$  is strictly longer than  $\sigma$ , which implies  $w_1 = \sigma u$ . Therefore  $w = \sigma u w'$ , for some  $w'$ . Hence  $\sigma w'$  is an element of  $\text{eq}(g, h)$  shorter than  $w$ , and thus an element of  $\sigma u^*v_0$ . Therefore  $w' \in u^*v_0$  and  $w \in \sigma u^*v_0$ , a contradiction.

We have seen that  $u$  is not a suffix of  $\sigma$ . Also  $\sigma$  cannot be a suffix of  $u$ , otherwise  $\sigma u \sigma^{-1} \in \text{Eq}(g, h)$  will contradict the minimality of  $\alpha$  and  $\beta$ . We can therefore define  $\rho$  as the longest common suffix of  $u$  and  $\sigma$ . The word  $\rho$  is not empty since  $z_h$  is a suffix of both  $g(u)$  and  $g(\sigma)$ , and  $\bar{g}$  is marked. Denote,  $\eta = v_0$ ,  $\zeta = \sigma \rho^{-1}$  and  $\mu = u \rho^{-1}$ .

Note that the word  $\tau = \rho \eta$  is the longest common suffix of  $\alpha$  and  $\beta$ . Lemma 21 applied to  $(\bar{g}, \bar{h})$  yields  $g(\rho \eta) = \underline{z}_h h(\rho \eta)$ . The verification of all claims in case (B) is now straightforward.  $\square$

Note that the previous lemma proves, in particular, Theorem 1(C).

The following lemma allows to suppose that both  $\bar{g}$  and  $\bar{h}$  are marked, which was the task of this section.

**Claim 4.** *Let  $(g, h)$  be a counterexample. Then there exists also a counterexample  $(g_1, h_1)$  such that both  $\bar{g}_1$  and  $\bar{h}_1$  are marked.*

*Proof.* Suppose that  $(g, h)$  is a counterexample. Then  $\bar{g}$  is marked by Claim 3. Suppose that  $\underline{z}_h \neq \varepsilon$  and define  $g_1$  and  $h_1$  by

$$g_1(u) = g(u), \quad h_1(u) = \underline{z}_h h(u)(\underline{z}_h)^{-1}.$$

It is not difficult to see that the morphism  $h_1$  is well defined, it is not marked while  $\bar{h}_1$  is marked. It remains to show that  $\text{Eq}(g_1, h_1)$  has rank at least two.

This is a consequence of the characterization presented in Lemma 30. (We shall use its notation.)

1. If the case (A) of Lemma 30 takes place, then

$$\{\tau\sigma\mu_a, \tau\sigma\mu_b\} \subset \text{Eq}(g_1, h_1).$$

$$\begin{array}{l} g : \\ h : \end{array} \boxed{\begin{array}{c} \tau\sigma \quad \boxed{z_{h_1} = z_h z_h} \quad \mu_x \end{array}}$$

2. If, on the other hand, we have the case (B) of the Lemma 30, then

$$\{\rho\mu, \rho\eta\zeta\} \subset \text{Eq}(g_1, h_1).$$

$$\begin{array}{l} g : \\ h : \end{array} \boxed{\begin{array}{c} \rho \quad \boxed{z_{h_1} = z_h z_h} \quad \mu \end{array}} \quad \begin{array}{l} g : \\ h : \end{array} \boxed{\begin{array}{c} \rho \quad \boxed{z_{h_1} = z_h z_h} \quad \eta\zeta \end{array}}$$

Definitions in Lemma 30 yield  $\text{pref}_1(\mu_a) \neq \text{pref}_1(\mu_b)$  and  $\text{pref}_1(\mu) \neq \text{pref}_1(\eta)$ , whence the equality set has in both cases rank at least two.  $\square$

## 9 Case: $\bar{g}$ and $\bar{h}$ marked.

From now on we shall suppose that both  $\bar{g}$  and  $\bar{h}$  are marked. Consider Lemma 30. It is easy to note that the case (A) of the lemma has to take place, and moreover, the word  $\tau$  is empty. Therefore

$$\text{eq}(g, h) = \{\sigma\mu_a, \sigma\mu_b\},$$

with  $\text{pref}_1(\mu_a) = a$ ,  $\text{pref}_1(\mu_b) = b$ , and  $\text{suff}_1(\mu_a) \neq \text{suff}_1(\mu_b)$ .

Note the following useful fact.

**Lemma 31.** *Let  $(g, h)$  be a counterexample such that  $\bar{g}$  and  $\bar{h}$  are marked. Put  $g_1 = \bar{g}$  and  $h_1 = \bar{h}_m$ . Then the pair  $(g_1, h_1)$  is again a counterexample such that  $\bar{g}_1$  and  $\bar{h}_1$  are marked, and*

$$\text{eq}(g_1, h_1) = \{\bar{\sigma}\bar{\mu}_a, \bar{\sigma}\bar{\mu}_b\}.$$

*Proof.* The verification is straightforward.  $\square$

In this section, we will also need to assume that the pair  $(g, h)$  is a *shortest counterexample*. That is,  $|\sigma\mu_a| + |\sigma\mu_b|$  is as small as possible. Shortest counterexample have the following important properties, which can be summarized as: there are no repeated overflows. The proof is similar to the proof of Lemma 29. If there is a repeated overflow, then we can decompose the counterexample into blocks, and find a shorter counterexample, namely the pair of successor morphisms. Since both  $\bar{g}$  and  $\bar{h}$  are marked, we will consider their blocks, which are easier to deal with.

**Lemma 32.** *Let  $(g, h)$  be a counterexample such that  $\bar{g}$  and  $\bar{h}$  are marked. Let two nonempty prefixes  $\sigma_1$  and  $\sigma_2$  of  $\sigma$  satisfy  $g(\sigma_1) = h(\sigma_2)$ . Then  $(g, h)$  is not a shortest counterexample.*

*Proof.* Lemma 19 applied to morphisms  $\bar{g}$  and  $\bar{h}$  implies that pairs  $(\overline{\sigma\mu_a}, \overline{\sigma\mu_a})$ ,  $(\overline{\sigma\mu_b}, \overline{\sigma\mu_b})$  and  $(\overline{\sigma_1}, \overline{\sigma_2})$  can be factorized into a sequence of pairs  $(\bar{e}, \bar{f})$  and  $(\bar{e}', \bar{f}')$  such that  $\bar{g}(\bar{e}) = \bar{h}(\bar{f})$  and  $\bar{g}(\bar{e}') = \bar{h}(\bar{f}')$ . Turning to reversals and defining  $g_1$  and  $h_1$  as in (5) we obtain words  $w, w' \in \text{Eq}(g_1, h_1)$  such that

$$g_1(w) = h_1(w) = \sigma\mu_a, \quad g_1(w') = h_1(w') = \sigma\mu_b.$$

Note also that  $\bar{g}_1$  and  $\bar{h}_1$  are marked by Lemma 24.

Since  $(\sigma_1, \sigma_2)$  is a prefix of both  $(\sigma\mu_a, \sigma\mu_a)$  and  $(\sigma\mu_b, \sigma\mu_b)$ , the words  $w$  and  $w'$  have a nonempty common prefix. From  $g \neq h$ , it is also easy to see that  $|w| + |w'| < |\sigma\mu_a| + |\sigma\mu_b|$ . Lemma 27 concludes the proof.  $\square$

**Lemma 33.** *Let  $(g, h)$  be a shortest counterexample such that  $\bar{g}$  and  $\bar{h}$  are marked. Let prefixes  $\sigma_1, \sigma_2, \sigma'_1$  and  $\sigma'_2$  of  $\sigma$  satisfy*

$$h(\sigma_2)^{-1}g(\sigma_1) = h(\sigma'_2)^{-1}g(\sigma'_1). \quad (19)$$

*Then  $\sigma_1 = \sigma'_1$  and  $\sigma_2 = \sigma'_2$ .*

Recall that we allow (19) to be an equality of two “negative” words if  $g(\sigma_1) < h(\sigma_2)$  and  $g(\sigma'_1) < h(\sigma'_2)$ .

*Proof.* Proceed by contradiction. Without loss of generality, we can suppose  $|\sigma_1| > |\sigma'_1|$  and  $|\sigma_2| > |\sigma'_2|$ . Let  $u_1$  be the longest common suffix of  $\sigma_1$  and  $\sigma'_1$ , and let  $u_2$  be the longest common suffix of  $\sigma_2$  and  $\sigma'_2$ . We want to show that

$$g(\sigma_1 u_1^{-1}) = h(\sigma_2 u_2^{-1}). \quad (20)$$

From (19) and  $g(\sigma\mu_a) = h(\sigma\mu_a)$ , we deduce

$$g(\sigma'_1 \sigma_1^{-1} \sigma\mu_a) = h(\sigma'_2 \sigma_2^{-1} \sigma\mu_a). \quad (21)$$

If  $u_1 = \sigma'_1$  and  $u_2 = \sigma'_2$ , then (20) follows from (21). Otherwise, we apply Lemma 16 to morphisms  $\bar{g}$  and  $\bar{h}$  (note that the role of  $\bar{g}$  and  $\bar{h}$  is interchangeable since both are marked), and to pairs

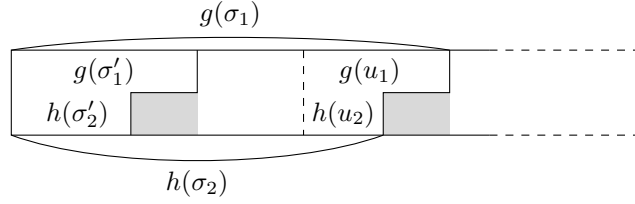
$$(\overline{\sigma'_1 \sigma_1^{-1} \sigma\mu_a}, \overline{\sigma'_2 \sigma_2^{-1} \sigma\mu_a}) \quad \text{and} \quad (\overline{\sigma\mu_a}, \overline{\sigma\mu_a})$$

to obtain

$$g(u_1 \sigma_1^{-1} \sigma\mu_a) = h(u_2 \sigma_2^{-1} \sigma\mu_a),$$

whence (20) follows too.

The rest is Lemma 32.



□

As a particular case, we point out the following corollary.

**Lemma 34.** *Let  $(g, h)$  be a shortest counterexample such that  $\bar{g}$  and  $\bar{h}$  marked. Let two prefixes  $\sigma_1$  and  $\sigma_2$  of  $\sigma$  satisfy  $g(\sigma_1) = h(\sigma_2)z_h$ . Then  $\sigma_1 = \sigma_2 = \sigma$ .*

### 9.1 The case: $\text{pref}_1(\sigma) = a$ or $\text{suff}_1(\sigma) = a$

In this subsection we show that the word  $\sigma$  of a counterexample cannot start nor end by the letter  $a$ .

Since  $|g(a)| > |h(a)|$  and  $\text{suff}_1(\mu_c) = a$  for some  $c \in A$ , we have

$$h(a) \in \text{suff}(g(a)). \quad (22)$$

**Claim 5.** *There is no counterexample such that both  $\bar{g}$  and  $\bar{h}$  are marked and  $\text{pref}_1(\sigma) = a$  or  $\text{suff}_1(\sigma) = a$ .*

*Proof.* Let first  $\text{pref}_1(\sigma) = a$ . As in the proof of Claim 3, we obtain that  $z_h$  and  $h(a)$  have a common primitive root, say  $t$ . From (22) we have that  $t$  is a suffix of  $g(a)$ , which yields a contradiction with Claim 2.

The case  $\text{suff}_1(\sigma) = a$  follows from the same considerations for morphisms  $\bar{g}$  and  $\bar{h}_m$  by Lemma 31. □

### 9.2 The case: $\text{pref}_1(\sigma) = \text{suff}_1(\sigma) = b$

In this subsection we shall suppose that  $(g, h)$  is a counterexample such that  $\bar{g}$  and  $\bar{h}$  are marked, and  $\text{pref}_1(\sigma) = \text{suff}_1(\sigma) = b$ . We shall restrict possible counterexamples to the case  $\mu_b \in b^+$ .

We first fix some notation.

*Convention 35.*

- Denote by  $\ell$  the maximal integer such that  $b^\ell$  is a prefix of  $\sigma$ .
- Denote by  $k$  the maximal integer such that  $b^k$  is a prefix of  $\mu_b\sigma$ .
- Denote by  $\ell'$  the maximal integer such that  $b^{\ell'}$  is a suffix of  $\sigma\mu_b$  or  $\sigma\mu_a$  (the one of the two equality words ending with  $b$ ).

- Denote by  $k'$  the maximal integer such that  $b^{k'}$  is a suffix of  $\sigma$ .

We make use of Lemma 31 and suppose that  $k' \geq \ell$ . In other words, we shall work either with  $(g, h)$  or with  $(\overline{g}, \overline{h_m})$  depending on whether  $\sigma$  has more  $b$ s in the front or in the rear.

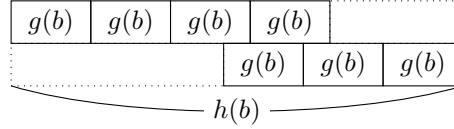
We first present some auxiliary lemmas.

**Lemma 36.** *The words  $g(b)$  and  $h(b)$  do not commute.*

*Proof.* Suppose, for a contradiction, that  $t$  is the common primitive root of  $g(b)$  and  $h(b)$ . Since  $|g(b)| < |h(b)|$ , we deduce, by  $g(\sigma) = h(\sigma)z_h$ , that the first occurrence of  $g(a)$  in  $g(\sigma)$  is comparable with  $t$ , a contradiction with  $g$  being marked.  $\square$

**Lemma 37.**  $|h(b)| > (\ell + \ell' - 1)|g(b)|$ .

*Proof.* The word  $h(b)$  is comparable with  $g(b)^\ell$  and suffix-comparable with  $g(b)^{\ell'}$ . If  $|h(b)| \leq (\ell + \ell' - 1)|g(b)|$ , then  $g(b)$  and  $h(b)$  commute by Lemma 7(B), a contradiction with Lemma 36.



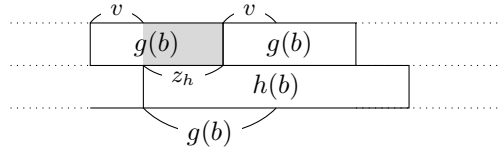
$\square$

**Lemma 38.**  $|z_h| > (\ell + k' - 1)|g(b)|$ .

*Proof.* The word  $z_h$  is comparable with  $g(b)^\ell$ , since  $z_h$  is comparable with  $h(b)$ , and  $g(b)^\ell$  is a prefix of  $h(b)$ . Also  $z_h$  is suffix-comparable with  $g(b)^{k'}$ , by  $g(\sigma) = h(\sigma)z_h$ .

First suppose that  $|z_h| \geq |g(b)|$ . Now, if  $|z_h| \leq (\ell + k' - 1)|g(b)|$ , then  $z_h$  and  $g(b)$  commute by Lemma 7(B), a contradiction with Claim 1.

Suppose now that  $z_h$  is shorter than  $g(b)$ . Recall that  $g(b)$  is a prefix of  $g(\mu_b)$ , prefix of  $h(b)$ , and a suffix of  $g(\sigma)$ . From  $g(\sigma) = h(\sigma)z_h$  and  $z_h g(\mu_b) = h(\mu_b)$  we deduce that there is a word  $v$  such that  $g(b) = vz_h$  and at the same time  $g(b) = z_h v$ . Again, the words  $g(b)$  and  $z_h$  commute, a contradiction.



$\square$

An important step in the proof is the following lemma which shows that  $g(a)$  cannot be too short.

**Lemma 39.**  $|g(ba)| > |h(b)|$ .

*Proof.* In this proof we shall consider occurrences of  $g(b)$ s and  $h(b)$ s in  $g(\sigma)$  and  $h(\sigma)$ , and their relative position. The idea is quite intuitive, but we give a more formal definition. Let  $i, j \leq |\sigma|_b$  be positive integers. Denote by  $u_i$  the prefix of  $\sigma$  such that also  $u_i b$  is a prefix of  $\sigma$ , and  $|u_i b|_b = i$ .

We say that the  $i$ th occurrence of  $g(b)$  in  $g(\sigma)$  *starts* within the  $j$ th occurrence of  $h(b)$  in  $h(\sigma)$ , if

$$|h(u_j)| \leq |g(u_i)| < |h(u_j b)|.$$

Similarly, we say that the  $i$ th occurrence of  $g(b)$  in  $g(\sigma)$  *ends* within the  $j$ th occurrence of  $h(b)$  in  $h(\sigma)$ , if

$$|h(u_j)| < |g(u_i b)| \leq |h(u_j b)|.$$

Lemma 38 implies that the last occurrence of  $g(b)$  in  $g(\sigma)$  both starts and ends outside  $h(\sigma)$ . Therefore, by the pigeon hole principle, there is an occurrence of  $h(b)$  in  $h(\sigma)$  such that no occurrence of  $g(b)$  in  $g(\sigma)$  starts within it. Similarly, there is an occurrence of  $h(b)$  in  $h(\sigma)$  within which no  $g(b)$  ends. From this it is easy to deduce that  $h(b)$  is a prefix of  $sg(a)^+$ , and a suffix of  $g(a)^+p$  where  $s$  is a suffix of  $g(b)$  or  $g(a)$ , and  $p$  is a prefix of  $g(b)$  or  $g(a)$ . Let  $t$  be the primitive root of  $g(a)$ .

By  $g(\sigma) = h(\sigma)z_h$ , the words  $h(b)$  and  $g(b^\ell a)$  are prefix comparable. Suppose that  $g(b^\ell t)$  is a prefix of  $h(b)$ . From the fact that  $g(b^\ell)g(a)$  is a prefix of  $sg(a)^+$  we deduce by Lemma 7(C) that  $\bar{g}$  is not marked, a contradiction. Similarly, we obtain a contradiction with  $g$  being marked, if  $tg(b)^{\ell'}$  is a suffix of  $h(b)$ . Therefore

$$|h(b)| < |g(b)^\ell t| \text{ and } |h(b)| < |tg(b)^{\ell'}| \quad (23)$$

and we are through if  $\ell = 1$ .

Suppose  $\ell \geq 2$ . Again by a pigeon hole principle, there are at least two occurrences of  $h(b)$  in  $g(\sigma)$  with no starting  $g(b)$ . Therefore  $h(b)$  is a prefix of  $s_1 g(a)^+$  and  $s_2 g(a)^+$ , where  $s_1$  and  $s_2$  are proper suffixes of  $g(b)$  or  $g(a)$ . Note that  $s_1$  and  $s_2$  are overflows in  $\sigma$ , whence  $s_1 \neq s_2$  by Lemma 33. Suppose, for a contradiction, that  $|g(ba)| \leq |h(b)|$ . From  $s_1 \neq s_2$ , it is then not difficult to deduce that  $g(a)$  overlaps nontrivially with  $g(a)^2$ , whence it is not primitive and  $|g(a)| \geq 2|t|$ . From this and from (23) we obtain

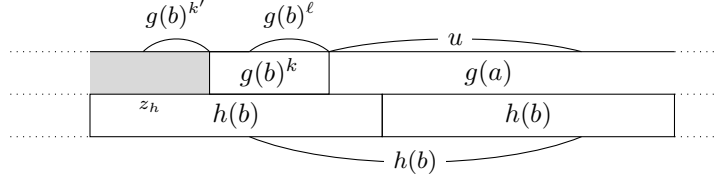
$$2|h(b)| < |g(b)^\ell t| + |g(b)^{\ell'} t| \leq (\ell + \ell')|g(b)| + |g(a)| \leq (\ell + \ell' - 1)|g(b)| + |h(b)|,$$

a contradiction with Lemma 37.  $\square$

We can now once more point out two commuting words.



**Lemma 40.** *The word  $h(b)$  commutes with  $z_h g(b)^{k-\ell}$ .*



*Proof.* Lemma 38 and the definition of  $k'$  implies that  $g(b)^{k'}$  is a suffix of  $z_h$ . The assumption  $k' \geq \ell$  guarantees that  $z_h g(b)^{k-\ell}$  is a well defined prefix of  $z_h g(b)^k$ .

Let  $u$  be the word  $g(b)^{-\ell} h(b)$ , which is a prefix of  $g(a)$  by Lemma 39. Since  $|h(b)| > |g(b)|$  and  $\ell \geq 1$ , we have

$$|h(b)^k z_h| > |z_h g(b)^{k-\ell} h(b)|.$$

The equality  $z_h g(\mu_b) = h(\mu_b)$  now implies that the word

$$z_h g(b)^k u = z_h g(b)^{k-\ell} h(b)$$

is a prefix of  $h(b)^+$  and thus  $z_h g(b)^{k-\ell}$  commutes with  $h(b)$  by Lemma 7(B).  $\square$

As a consequence, we have the claim of this section.

**Claim 6.** *If  $(g, h)$  is a shortest counterexample such that  $\text{pref}_1(\sigma) = \text{suff}_1(\sigma) = b$ , then  $\mu_b = b^{k-\ell}$ .*

*Proof.* Let  $t$  be the common primitive root of words  $h(b)$  and  $z_h g(b)^{k-\ell}$ . Recall that, by (1), the maximal  $t$ -prefix of any  $h(au)z_h$  is  $z_h$ . We deduce the following.

- The maximal  $t$ -prefix of  $h(\sigma)z_h = g(\sigma)$  is  $h(b)^\ell z_h$ .
- The maximal  $t$ -prefix of  $h(\mu_b \sigma)z_h = z_h g(\mu_b \sigma) = z_h g(b)^{k-\ell} g(b^{\ell-k} \mu_b \sigma)$  is  $h(b)^k z_h$ , which implies that the maximal  $t$ -prefix of  $g(b^{\ell-k} \mu_b \sigma)$  is

$$(z_h g(b)^{k-\ell})^{-1} h(b)^k z_h = h(b)^k g(b)^{\ell-k}.$$

Since  $\sigma$  contains  $a$ , both maximal  $t$ -prefixes mentioned above are proper.

Let first  $h(b)^k g(b)^{\ell-k} \neq h(b)^\ell z_h$ , and put  $v = \sigma \wedge b^{\ell-k} \mu_b \sigma$ .

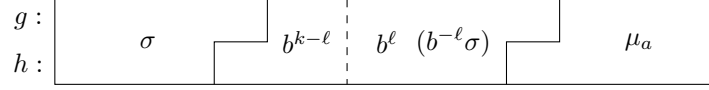
If  $|h(b)^k g(b)^{\ell-k}| > |h(b)^\ell z_h|$ , then  $g(v) = h(b)^\ell z_h$ , by Lemma 8, a contradiction with Lemma 34.

On the other hand,  $|h(b)^k g(b)^{\ell-k}| < |h(b)^\ell z_h|$  implies  $k < \ell$ , and Lemma 8 yields  $g(v) = h(b)^k g(b)^{\ell-k}$  and  $g(v b^{k-\ell}) = h(b)^k$ , a contradiction with Lemma 32.

It remains that  $h(b)^k g(b)^{\ell-k} = h(b)^\ell z_h$ , which implies  $h(b)^{k-\ell} = z_h g(b)^{k-\ell}$  and  $\mu_b = b^{k-\ell}$ .  $\square$

### 9.3 The case: $\text{pref}_1(\sigma) = \text{suff}_1(\sigma) = b$ and $\mu_b = b^{k-\ell}$

This last case is most difficult because it in a way compresses two places we use for the analysis into one, namely the beginning of  $\sigma$  and the beginning of  $\mu_b$ .



Therefore, we have to employ a more detailed analysis of  $\mu_a$ .

**Claim 7.** *There is no shortest counterexample with  $\text{pref}_1(\sigma) = \text{suff}_1(\sigma) = b$ .*

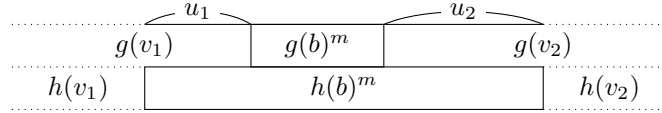
*Proof.* Claim 6 implies  $\text{pref}_1(\mu_b) = \text{suff}_1(\mu_b) = b$  whence

$$\text{pref}_1(\mu_a) = \text{suff}_1(\mu_a) = a.$$

Let  $v_1$  be the longest prefix of  $\mu_a$  ending with  $a$  and satisfying  $|z_h g(v_1)| > h(v_1)$ . It follows that  $\mu_a = v_1 b^m v_2$  where  $m > 0$ ,  $\text{suff}_1(v_1) = \text{pref}_1(v_2) = a$  and

$$|z_h g(v_1)| > |h(v_1)|, \quad |g(v_2)| > |h(v_2)|.$$

Denote  $u_1 = h(v_1)^{-1} z_h g(v_1)$  and  $u_2 = g(v_2) h(v_2)^{-1}$ .



From  $|h(b)| > |g(b)|$  and  $z_h g(b^{k-\ell}) = h(b^{k-\ell})$  we obtain

$$|h(b)| \leq |z_h g(b)|. \quad (24)$$

Let now  $y$  be the prefix of  $g(ba)$  of length  $h(b)$  and let  $x$  be the word such that  $xg(b)^{\ell'} = h(b)^{\ell'}$ . From (24) we deduce that  $u_1 g(b)^{m-1} y$  is a prefix of  $h(b)^+$ . Also  $xg(b)^{\ell'+\ell-1} y$  is a prefix of  $h(b)^+$ . Lemma 7(E) now implies that  $u_1 g(b)^{m-1}$  and  $xg(b)^{\ell'+\ell-1}$  are suffix comparable. Since  $\bar{g}$  is marked and both  $x$  and  $u_1$  are suffix comparable with  $g(a)$ , we deduce  $m = \ell + \ell'$ .

From Lemma 39, we obtain that  $h(b)$  is a prefix of  $g(b)^\ell g(a)$  whence the word  $u_1 g(b)^{\ell'} h(b)$  is a prefix of  $h(b)^m z_h$ . Lemma 7(B) now implies that  $u_1 g(b)^{\ell'}$  commutes with  $h(b)$ .

Note that  $u_1 g(b)^{\ell'}$  is the maximal  $h(b)$ -suffix of  $g(\sigma v_1 b^{\ell'})$  and  $xg(b)^{\ell'}$  is the maximal  $h(b)$ -suffix of  $g(\sigma \mu_b)$ . Minimality of  $\sigma \mu_a$  implies that

$$u_1 g(b)^{\ell'} \neq h(b)^{\ell'} = xg(b)^{\ell'}.$$

Let  $v$  be the longest common suffix of  $\sigma \mu_b$  and  $\sigma v_1 b^{\ell'}$ . We apply Lemma 8 to  $\bar{g}$  and obtain that  $g(v)$ , which is the longest common suffix of  $g(\sigma \mu_b)$  and  $g(\sigma v_1 b^{\ell'})$ , is equal either to  $u_1 g(b)^{\ell'}$  or to  $xg(b)^{\ell'}$ . In both cases,  $g(v)$  commutes with  $h(b)$ ; let  $t$  be their common primitive root.

Since  $\bar{g}$  is marked, the maximal  $t$ -suffix of  $g(\sigma\mu_b)$  is  $g(u)$  where  $u$  is the maximal  $v$ -suffix of  $\sigma\mu_b$ . Since  $\bar{h}$  is marked, the maximal  $t$ -suffix of  $h(\sigma\mu_b)$  is  $h(b^{\ell'})$ . Therefore  $g(\sigma\mu_b u^{-1}) = h(\sigma\mu_b b^{-\ell'}) = h(\sigma b^{-k'})$ , where  $\sigma\mu_b u^{-1}$  is a prefix of  $\sigma$  since  $h(\sigma b^{-k'})$  is a prefix of  $h(\sigma)$ . Hence  $(g, h)$  is not a shortest counterexample by Lemma 32.  $\square$

This concludes the proof that there is no counterexample. By Lemma 27, two minimal elements  $\alpha$  and  $\beta$  of  $\text{Eq}(g, h)$  cannot start with the same letter if  $g$  and  $h$  are both non-periodic. Clearly, also  $\bar{g}$  and  $\bar{h}$  are non-periodic and  $\bar{\alpha}, \bar{\beta}$  are minimal elements of  $\text{Eq}(\bar{g}, \bar{h})$ . Theorem 2 is proved.

## 10 Test set

In this section we show that each binary language has a test set of cardinality at most two. The result is a consequence of Theorem 1 and Theorem 2.

*Test set* of a language  $L \subset \Sigma^*$  is defined as a subset  $T$  of  $L$  such that the agreement of two morphisms on the language  $T$  guarantees their agreement on  $L$ . Formally, for any two morphisms  $g$  and  $h$  defined on  $\Sigma^*$

$$(\forall u \in T) (g(u) = h(u)) \Rightarrow (\forall v \in L) (g(v) = h(v)).$$

The *ratio* of a word  $u \in A^+$  is denoted by  $r(u)$  and defined by

$$r(u) = \frac{|u|_a}{|u|_b}.$$

If  $|u|_b = 0$ , then  $r(u) = \infty$ . A word  $u$  is said to be *ratio-primitive* if no proper prefix of  $u$  has the same ratio as  $u$ .

It is not difficult to see that each nonempty word  $u$  has a unique factorization  $u = u_1 \dots u_k$  where each  $u_i$  is a nonempty ratio-primitive word such that  $r(u_i) = r(u)$ . We call it the *ratio-primitive factorization* of  $u$ . Let  $R(L)$  denote the set of all ratio-primitive words  $u$  such that  $u$  occurs in the ratio-primitive factorization of at least one word in  $L$ .

**Lemma 41.** *If  $|g(a)| \neq |h(a)|$ , then  $|g(u)| = |h(u)|$ ,  $u \in A^+$ , if and only if*

$$r(u) = \frac{|h(b)| - |g(b)|}{|g(a)| - |h(a)|}.$$

*If  $|g(a)| = |h(a)|$  and  $|g(b)| \neq |h(b)|$ , then  $|g(u)| = |h(u)|$ ,  $u \in A^+$ , if and only if  $r(u) = \infty$ .*

*Proof.* Follows directly from

$$|g(u)| = |u|_a \cdot |g(a)| + |u|_b \cdot |g(b)| \quad \text{and} \quad |h(u)| = |u|_a \cdot |h(a)| + |u|_b \cdot |h(b)|.$$

$\square$

An immediate corollary is the following fact.

**Lemma 42.** *Binary morphisms  $g$  and  $h$  agree on  $L$  if and only if they agree on  $R(L)$ .*

Here is one more observation.

**Lemma 43.** *If  $g(u) = h(u)$  and  $g(v) = h(v)$ , with  $u, v \in A^+$  and  $r(u) \neq r(v)$ , then  $g = h$ .*

*Proof.* Since  $r(u) \neq r(v)$ , the word  $uv$  contains both letters  $a$  and  $b$ . Lemma 41 implies  $|g(a)| = |h(a)|$  and  $|g(b)| = |h(b)|$  whence  $g = h$ .  $\square$

We can now prove the main claim.

**Theorem 44.** *Let  $L \subset A^*$  be a language. Then  $L$  possesses a test set of cardinality at most two.*

*Proof.* If  $L$  contains words  $u$  and  $v$  with different ratios, then  $T = \{u, v\}$  is a test set of  $L$  by Lemma 43.

Suppose that all words in  $L$  have the same ratio. We first find a test set  $T_R$  of cardinality at most two for  $R(L)$ . If  $R(L)$  has cardinality at least three, let  $T_R = \{u, v\}$  where  $u, v \in R(L)$  and  $\text{pref}_1(u) = \text{pref}_1(v)$ .

Let  $g$  and  $h$  be morphisms such that  $g \neq h$ ,  $g(u) = h(u)$  and  $g(v) = h(v)$ . Since  $u$  and  $v$  are ratio-primitive, Lemma 41 implies that  $u$  and  $v$  are minimal elements of  $\text{Eq}(g, h)$ . Therefore both morphisms are periodic by Theorem 1(B) and Theorem 2. By Theorem 1(A), we have  $R(L) \subseteq \text{Eq}(g, h)$ .

Let now  $T$  be a subset of  $L$  such that  $R(T) = T_R$ . Clearly,  $T$  can be chosen such that its cardinality is at most two. Lemma 42 concludes the proof.  $\square$

## References

- [1] J. Berstel, D. Perrin, J.-F. Perrot, and A. Restivo. Sur le théorème du défaut. *J. Algebra*, 60(1):169–180, 1979.
- [2] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages, Vol. 1*, pages 329–438. Springer, Berlin, 1997.
- [3] K. Čulík, II and J. Karhumäki. On the equality sets for homomorphisms on free monoids with two generators. *RAIRO Inform. Théor.*, 14(4):349–369, 1980.
- [4] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. On binary equality sets and a solution to the test set conjecture in the binary case. *J. Algebra*, 85(1):76–85, 1983.
- [5] A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg. The (generalized) Post correspondence problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.*, 21(2):119–144, 1982.

- [6] T. Harju and J. Karhumäki. Morphisms. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages, Vol. 1*, pages 439–510. Springer, Berlin, 1997.
- [7] Š. Holub. A unique structure of two-generated binary equality sets. In *Developments in language theory*, volume 2450 of *Lecture Notes in Comput. Sci.*, pages 245–257. Springer, Berlin, 2003.
- [8] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1983.
- [9] A. Salomaa. Equality set for homomorphisms of free monoids. *Acta Cybernet.*, 4(1):127–139, 1978/79.
- [10] Štěpán Holub. *Equations in free monoids*. PhD thesis, Charles University, Prague, 2000.